

# GEROS Evaluation Quality Assurance Tool

Version: September 2021

## Summative Evaluation of the UN Joint Programme on Girls Education

REPORT RATING SUMMARY			
<b>Overall Rating</b>	78%	Satisfactory	
●●●●●	Exceptional (96% - 100%)	5	
●●●●	Highly Satisfactory (87.5% - 95.99%)	4	
●●●	<b>Satisfactory (62.5% - 87.49%)</b>	3	Meets UNICEF/UNEG standards for evaluation reports and decision makers may use the evaluation with confidence
●●	Fair (35% - 62.49%)	2	
●	Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS	
<b>Title of the evaluation report</b>	Summative Evaluation of the UN Joint Programme on Girls Education
<b>Report sequence number</b>	Malawi 63-2024-20337
<b>Region</b>	ESAR
<b>Year of report</b>	2024
<b>Office</b>	Malawi
<b>Coverage (countries)</b>	Malawi
<b>ToRs present</b>	Yes
<b>Date of review (dd/mm/yyyy)</b>	February 10, 2025
<b>Name of review firm</b>	IOD PARC

CLASSIFICATION OF EVALUATION REPORT	
<b>Management of evaluation (Managerial control and oversight of evaluation)</b>	Jointly managed with one or more UN agencies
<b>Unicef goal areas (Alignment with strategic plan priorities)</b>	
Every child survives and thrives	No
Every child learns	Yes
Every child is protected from violence and exploitation	No
Every child lives in a safe and clean environment	No
Every child has an equitable chance in life	No
Gender equality (cross-cutting)	Yes
Humanitarian action (cross-cutting)	No
<b>Evaluation object</b>	Programme
<b>Evaluation type</b>	Summative
<b>Evaluation strategy</b>	Mixed methods
<b>Evaluation design (primary method used)</b>	Quasi-experimental
<b>Evaluation level</b>	Impact
<b>Geographic scope</b>	National
<b>Primary SDG(s) covered (number)</b>	4

**EQA Summary:** The rater will provide top line issues for this evaluation relevant for feedback to senior management (positive and negative), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater will highlight best practice/added value elements and the level of complexity of the evaluation.

This is a good evaluation report that is well written and provides a robust impact analysis of the Joint Programme for Girls' Education in Malawi programme over three phases between 2014 and 2024. Some top-line issues are presented below:

- The evaluation benefits from a strong methodology that responds well to the evaluation scope and objectives through a combination of qualitative and quantitative data collection and analysis approaches that allows for an effective examination of factors influencing the overall findings.
- The gender focus and integration of disability considerations is particularly welcome, and the evaluation process involved efforts to identify and include multiple stakeholder perspectives including those of children, adolescents, parents and community leaders as well as UN, national and district level government representatives, school teachers and principals.
- The detailed methodological focus on outcome to impact analysis is a particular strength, especially in its adoption of mixed methods, sources, and a statistically valid sampling approach. While the specific details may not be applicable to all evaluations, the attention to methodological rigour and triangulation to support the impact analysis and examination of contributions represents good practice that can be applied in other evaluations.
- The report is supplemented by Annexes that provide a useful additional resource to the evaluation methodology and findings. However, at 100 pages the evaluation is rather long. Opportunities to shorten the report include presenting some of the methodological detail in an Annex, and shortening the findings by the wider use of graphics, tables and boxes.

**Recommendations for Improvement:** The rater will identify topline recommendations to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, resources will be cited to assist evaluation managers in overseeing future evaluations.

Two broad top-line recommendations can be identified that would potentially help structure and present this and similar evaluation terms of reference and reports:

- **Focusing the evaluation enquiry on key strategic questions:** Thirty-three evaluation questions are described and presented against the OECD/DAC evaluation criteria of relevance, coherence, efficiency, effectiveness, sustainability, and impact. Although this formulation is acceptable and is present across multiple UNICEF and other agency evaluations, 33 questions arguably pushes the evaluation into an examination of specific areas of detail rather than higher level, more strategic lines of enquiry. Some questions might also be considered indicators. For example, under the Impact criterion, the question "How many children, including adolescents, girls and boys, and children with disabilities, have benefitted (and in what way) so far?" might have better considered an indicator for a higher level question on gender and inclusion. An alternative approach to structuring the enquiry around a shorter list of critical impact questions could have been considered in the TOR and evaluation.
- **Structuring the presentation of conclusions and recommendations:** The current presentation of conclusions and recommendations is a comparative weaknesses in the report. Both sections would have benefited from a little more consideration of higher level, strategic utility, involving a stronger focus on the evaluation objectives and description of actionable recommendations with clear roles and responsibilities for leadership and support. To achieve this, the evaluation report might have considered a more purposeful structure based around the 5 evaluation objectives rather than the rather cumbersome and abstract evaluation criteria. As an example, an objectives-focused structure would have allowed the evaluation report to structure and present the conclusions and recommendations around: (i) summarising the achievement of program objectives and outcomes; (ii) exploring areas leading to long-term impacts; (iii) identifying examples of best practice and failures; (iv) drawing out key management considerations that require greater attention among decision-makers (e.g. resourcing, fund allocation and management, and areas for JPGE modifications); and, (v) identifying critical accountability and transparency considerations that will be required to support future implementation (e.g. in district coordination and joint programme monitoring). This would have allowed the conclusions to act as a link that draws on and adds value to the findings and points toward the recommendations, and avoid the overly granular presentation of recommendations.

SECTION A: EXECUTIVE SUMMARY (weight 5%)		83%	Comments on Rating
<b>Question 1.</b>	Can the executive summary inform decision-making?		
i	Is clearly presented, serves as a standalone document useful for informing decision making, and is of relevant conciseness and depth for key users (Maximum of 5 pages unless otherwise specified in ToR).	Partially	The executive summary is clearly presented. At 6 pages it is a little longer the recommended page limit but is well written with relevant information that should support users and decision-makers of the Joint Programme on Girls' Education (JPGE)
ii	Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key conclusions on findings, lessons learned if requested, key recommendations) as per the ToR.	Yes	All relevant sections are included. The Findings are appropriately presented against the evaluation conclusions which are structured in line with the evaluation criteria.
iii	Includes all significant information needed to understand the intervention and the evaluation AND does not introduce new information from what is presented in the rest of the report.	Yes	The executive summary includes relevant information needed to understand the JPGE and its overall results. It does not introduce additional information to that presented in the main report.

SECTION B: BACKGROUND (weight 5%)		86%	Comments on Rating
<b>Question 2.</b>	Is the object of the evaluation clearly described?		
	i Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status.	Partially	The evaluation report (ER) includes a good description of the JPGE including the four districts covered by the programme, summary of the three implementation phases and their respective timelines over the period 2014-2024, as well as relevant national policies. Overall funding for the programme is included in the TOR, inception report (IR) and executive summary, but is not summarised for the three programme phases in the main report. A simple table with funding by year would have been sufficient.
	ii Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) by type (i.e., institutions/organizations; communities; individuals...), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cities, sub-regions...) and in terms of numbers reached, with disaggregation by gender, age, disability... (as appropriate to the purpose of the evaluation).	Yes	The different stakeholders are identified and number of beneficiaries is provided, disaggregated by core programme activity. A summary of beneficiaries and duty holders is included in the background. Beneficiary data is appropriately disaggregated by sex and age in the findings analysis of effectiveness and impacts, as well as for disability. The report also adopts appropriate rights based language.
<b>Question 3.</b>	Is the context of the intervention clearly described?		
	i Clear and relevant description of the context of the intervention (i.e. relevant policy, socio-economic, political, cultural, power/privilege, institutional, international factors) and how context relates to the implementation of the intervention.	Yes	Both the ER and IR include a good summary of relevant information. This includes an overview of primary and secondary education in Malawi, the context of girls' education, policy environment, summary of key stakeholders and institutions, and review of progress towards girls' education in Malawi.
	ii Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	Partially	Although a clear reference to SDG 4 (education) is provided, and indirect contributions to SDGs 2 (zero hunger), 3 (health and wellbeing), and 17 (partnerships) are identified, the ER does not identify or review JPGE contributions to SDG 4 targets or indicators.
	iii Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Yes	The ER and IR include good summaries of the education status and access to education services among Malawi school-children including girls, boys and children with disabilities. The evaluation background also reviews the state of social support for education in Malawi.
<b>Question 4.</b>	Are key stakeholders, their relationships and contributions clearly identified?		
	i Identification of implementing agency(ies), development partners, right holders, and additional duty bearers and other stakeholders; and of linkages between them (e.g., stakeholder map) (if relevant).	Yes	The JPGE implementing agencies are clearly identified as part of the summary of stakeholders (actors) in section 2.1 and the evaluation audience (3.1), as well as the evaluation methodology which uses key informant interview guides tailored to the different stakeholders identified in the evaluation background.
	ii Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Yes	The roles and contributions of stakeholders are summarised. Given the ER page count quite high, this is an example of an area where a table (and/or annex if necessary) could have been used to summarise and present stakeholder information.
<b>SECTION C: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)</b>		100%	Comments on Rating
<b>Question 5.</b>	Is the purpose of the evaluation clearly described?		
	i Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Yes	A clear summary of the evaluation purpose is provided in the TOR, IR and ER, alongside intended users.
<b>Question 6.</b>	Are the objectives and scope of the evaluation clear and realistic?		
	i Clear and complete description of what the evaluation seeks to achieve by the end of the process with reference to any changes made to the objectives included in the ToR (if applicable).	Yes	A clear description of the evaluation objectives is provided with sufficient detail provided to understand what the evaluation was commissioned to achieve. The 5 main objectives included: to, (i) Measure the achievement of program objectives and outcomes; (ii) Assess long-term impacts; (iii) Identify best practices, achievements, and failures; (iv) Inform decision-makers (vis a vis resource mobilisation and allocation, potential continuation, and/or modifications required to the JPGE design); and to, (v) Support accountability and transparency. No changes were made to the evaluation objectives between the TOR and final evaluation.
	ii Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	The evaluation scope is adequately summarised including its programme focus, analysis of all three JPGE phases (2014-2024); review of government education indicators, and geographic focus. No changes were made to the scope described in the TOR and IR.
<b>Question 7.</b>	Is the theory of change, results chain or logic well articulated?		
	i Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	Yes	The intended results of the JPGE are adequately described in Section 2 (programme description) and cover results across 5 core programme areas. These include: (i) the provision of Home Grown School Feeding (HGSF); (ii) improving the learning environment for children; (iii) training teachers; (iv) providing safe spaces for out of school girls (supporting foundational literacy, community based education (CBE) and vocational skills); and, (v) providing adolescents and young people with Sexual and Reproductive Health and Rights (SRHR) support (school, out of school, mobile clinics, youth friendly corners). Annex 3 presents a comprehensive list of indicators used by agencies over the three programme phases which are briefly summarised in the narrative.
	ii Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logic model, theory of change, evaluation matrix).	Yes	The causal relationship between the activity areas described under 7.i. and three programme outcomes is adequately described in the narrative and further presented as a theory of change in graphic form in Annex 4. The 3 outcomes include, 1. Increased access to quality and inclusive education by girls, boys, and adolescents (especially the most vulnerable) delivered through integrated services in a safe and gender transformative school, that enhances learning outcomes. 2. Increased access to complementary alternative learning and life skills and integrated services by girls and boys out of school. And, 3. Increased investment and support for education, life skills, health and nutrition of children and adolescents in and out-of-school by communities, parents, and education stakeholders.
	iii For theory-based evaluations, the theory of change or results framework is assessed, and if requested in the ToR, it is reformulated/improved by the evaluators.	Yes	The evaluation reviewed the theory of change presented in Annex 4 in its description of the JPGE, and development of the evaluation methodology.
<b>SECTION D: EVALUATION DESIGN AND METHODOLOGY (weight 20%)</b>		82%	Comments on Rating

<b>Question 8.</b>	Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? <i>UNICEF evaluation standards refer to the OECD/DAC criteria - Relevance; Coherence; Effectiveness; Efficiency; Sustainability; Impact (not all are necessarily relevant for all evaluations). Evaluations should also consider equity and leaving no-one behind, gender and human rights based approach (these can be mainstreamed into other criteria). Humanitarian evaluations should also consider Coverage; Connectedness; Coordination; Protection; Security.</i>		
	i Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Partially	Thirty-three evaluation questions are described and presented against the OECD/DAC evaluation criteria of relevance, coherence, efficiency, effectiveness, sustainability, and impact. Both the questions and criteria are adequately aligned and support the examination of evaluation objectives. One question is added to the 32 questions asked the TOR. The narrative includes a brief but adequate explanation for this. Although this formulation of evaluation questions is acceptable and is present across multiple UNICEF and other agency evaluations, 33 questions arguably pushes the evaluation into an examination of specific areas of detail rather than higher level, more strategic lines of enquiry.  Some questions might also be considered indicators. For example, under the Impact criteria, one question asks - "How many children, including adolescents, girls and boys, and children with disabilities, have benefitted (and in what way) so far?" This might have better considered an indicator for a higher level question on gender and inclusion. An alternative approach to structuring the enquiry around a shorter list of critical impact questions could have been considered in the TOR and evaluation.
	ii In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions and/or other processes from which the analysis can be based and conclusions drawn.	Partially	Annex 2 provides an evaluation matrix (called the 'evaluation framework') that summarises the evaluation dimensions, questions, indicators and methods used. The evaluation dimensions column adds little to what is already known from the criteria and questions. Dividing the methods column into separate columns for data collection and data analysis would have allowed the matrix to more directly respond to the methodology section.
<b>Question 9.</b>	Does the report specify adequate methods for data collection, analysis, and sampling?		
	i Evaluation design and set of methods are relevant and adequately robust for the evaluation's purpose, objectives and scope; and are fully and clearly described.	Yes	The evaluation methods for data collection and analysis are clearly described and provide a relevant and robust basis for data triangulation and the analysis of the evaluation questions and criteria that relate directly to the evaluation purpose and scope.
	ii Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.	Yes	A range of mixed data sources are fully described as part of a comprehensive presentation of the evaluation methodology and design. While this level of detail is appreciated, the section on data sources could have been shortened by providing more of a summary in the main narrative, and moving the detailed description of the evaluation approach, use of different sources, and justification to an Annex.
	iii Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Yes	The evaluation presents a complete and thorough set of indicators and sources across the implementing agencies. Mixed sources and methods for data collection are fully described under the sampling strategy. The narrative includes a comprehensive description of the qualitative and quantitative sampling strategies, including a statistical review of the appropriate target sampling size for the different impact surveys with head teachers, learners, parents and out-of-school parties.
	iv Clear and complete description of the methods of analysis.	Yes	The ER presents a complete and comprehensive description of the methods of analysis. These include a quasi-experimental longitudinal analysis of government data alongside a counterfactual sample (targeted, non-targeted schools); contribution analysis covering outcomes, causal pathways and contextual drivers/barriers; implementation analysis across stakeholders; relevance analysis; and use of applied statistical analysis across impact surveys.
	v Methodology allows for drawing causal connections between outputs and expected outcomes.	Yes	The use of mixed data sources and methods of analysis gives strong emphasis to contribution analysis and the examination of causal connections between activities, outputs, outcomes and impacts.
	vi Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).	Yes	The ER includes a good description of the evaluation limitations including data gaps in the government education management information system (EMIS) and learner survey. Section 3.3 (Limitations) explores data quality considerations, mitigation and appropriate judgements as to the extent to which the evaluation limitations posed a particular concern in specific analysis areas.
<b>Question 10.</b>	Are ethical issues and considerations described? The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include:		
	i Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or the principles in the 2020 revised UNEG Ethical Guidelines (integrity, accountability, respect, beneficence).	Yes	A comprehensive list of evaluation principles and guidance, and review of the evaluation adherence to UNEG, UNICEF and the evaluating agency's ethical considerations is provided. Annex 9 provides a letter outlining UNICEF's Ethics Review Board findings for the evaluation (reference, HML IRB Review 978MALW24).
	ii Description of ethical safeguards for participants appropriate for the issues relevant to methodology and how they are applied (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm). For those cases where the evaluation involves interviewing children, explicit reference is made to the UNICEF procedures for Ethical Research Involving Children.	Yes	Evidence of due consideration to ethical safeguards and the engagement of children. This includes with respect to informed consent, anonymity, safety and protection, evaluator training, adherence to UN standards compliance and data protection and UNICEF procedures for ethical research involving children, and team member selection for interviews and group discussions.
<b>Question 11.</b>	Does the evaluation incorporate innovative practice that adds value to the evaluation process?		
	i Innovation practice is used to improve the quality of evaluation process. This could be evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong child rights focus), or ways of sharing of evaluation results.	No	There are no entirely new or innovative approaches per se.
<b>SECTION E: EVALUATION FINDINGS (weight 25%)</b>		92%	Comments on Rating
<b>Question 12.</b>	Do the findings clearly address all evaluation objectives and scope?		

	i Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions, sub-questions and criteria.	Yes	The findings respond systematically to the evaluation questions. The sequence of the results sections is presented in line with evaluation criteria and their specific evaluation questions in line with the methodology and evaluation framework (matrix). Each of the findings is clearly presented with a headline finding statement followed by sufficient levels of evidence to explain the analysis. The presentation of the evaluation question and sub-questions above each of the findings statements is helpful. Two areas could be improved: a short explanatory analysis could be provided (in bold) with each of the findings statements such that each can be read as a stand-alone summary; and the length of presentation of the findings could be reduced by reducing the reliance on a narrative presentation and using more figures, tables and boxes to summarise the key points made and/or representative examples.
	ii Explicit use of the intervention's results framework/ToC in the formulation of the findings.	Yes	The findings include explicit reference of the JPGE results framework allowing the reader to build a picture of overall contributions and results.
<b>Question 13.</b>	Are evaluation findings derived from the conscientious, explicit and judicious use of the best available, objective, reliable and valid data and by accurate quantitative and qualitative analysis of evidence.		
	i Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident through the use of multiple data sources.	Yes	The findings involve credible, relevant and robust data analysis and triangulation and attend to the specific questions, sources and analytical methods outlined in the methodology and evaluation matrix. The appropriate use of interview quotes provides a rich source of supporting analysis, such as in the review of contributions of iron and folic acid supplements to girls' attendance in school (finding 37).
	ii Findings are clearly supported by, and respond to, the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.	Yes	The findings follow the structure of the evaluation criteria and questions, and demonstrate clear use of different data and analysis sources that respond to the indicators of the JPGE and evaluation matrix. Both positive and negative examples are provided. For example, the analysis of efficiency identifies how the timeliness of different interventions varied and often failed to meet planned implementation schedules leading to delays in teacher training sessions, infrastructure development projects, and community-level advocacy campaigns. Logistical, bureaucratic and fund disbursement constraints are identified as causal factors. The ER also reports unintended effects such as how the construction of girls' changing rooms was the reinforcement of menstruation stigma surrounding leading to some girls shunning to use the change rooms.
	iii Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings also analyse the logical chain (progression -or not- from implementation to results).	Yes	Causal factors, possible explanations and limits to the analysis are included. For example Finding 4 examines how the JPGE responded to the impacts of the COVID pandemic and a cholera outbreak on schools, and how closures were mitigated, such as through the use of take-home rations and activated existing WASH interventions to help mitigate health concerns. Similarly, Finding 39 described how "Experiences of violence remain prevalent in schools, with treatment schools reporting higher incidences of physical and sexual violence [but]... may be linked to heightened awareness of what constitutes violation and improved reporting mechanisms introduced by the JPGE program, which [has] encouraged more learners to disclose incidents of violence".
<b>Question 14.</b>	Does the evaluation assess and use the intervention's Results Based Management elements?		
	i Assessment of the adequacy of the intervention's monitoring system (including completeness and appropriateness of results/performance framework - including vertical and horizontal logic, M&E tools and their usage) to support decision-making.	Partially	Multiple findings identify weaknesses in programme monitoring (e.g. findings 17, 27, 39, 40, and 44) and the evaluation lesson 9 (section 5.2) responds reflects on the need for the JPGE to improve its overall approach to monitoring. While this is appropriate, the primary evaluation focus on understanding impacts has left a gap in the depth of analysis in this area. A little more consideration of which critical monitoring factors the UN agencies, and central and district government entities should address if they are to achieve a more integrated approach to monitoring would have been helpful. For example, is this more about data collection, analysis, capacity strengthening or data integration for decision-makers? Inclusion of a dedicated subquestion or indicator that explored the quality and effectiveness of monitoring as part of the analysis of inter-agency collaboration might have helped in this area.
<b>SECTION F:</b>	<b>EVALUATION CONCLUSIONS &amp; LESSONS LEARNED (weight 10%)</b>	<b>50%</b>	<b>Comments on Rating</b>
<b>Question 15.</b>	Do the conclusions clearly present an objective overall assessment of the intervention?		
	i Conclusions are clearly formulated and reflect the purpose and objectives of the evaluation. They are sufficiently forward looking (if a formative evaluation or if the implementation is expected to continue or have additional phase).	Partially	While the conclusions present many of the critical elements identified in the findings, the section is quite limited in its strategic insight as to the potential continuation of the JPGE. This is not helped by the use of a single, short narrative of just one and a half pages that lacks sub-titles to signpost the reader to the critical areas of analysis. This makes it quite difficult to relate the conclusions back to the evaluation's 59 findings, in order to understand what was or was not important when reflecting on the evaluation objectives. An alternative presentation might have tackled this directly by formulating and presenting the conclusions against each of the evaluation questions in turn, providing the reader with a headline conclusion to each question, and following it with the critical supporting evidence.
	ii Conclusions are derived appropriately from findings, and present a picture of the strengths and limitations of the intervention that adds insight and analysis beyond the findings.	Partially	While the conclusions do derive from the findings, they remain a bit passive. As a result they do not present a clear picture of the different strengths and weaknesses of the JPGE or add specific insights and strategic analysis. For example, one critical area is summarised in the single sentence, "Capacity to continue the multi-sectoral programme is still weak within the government despite willingness to continue based on the benefits observed". This begs multiple questions as to which areas are particularly problematic, at what level(s) of government, what should be done, whether there are nevertheless areas that do show strengths that can be built on, and the future role of the UN agencies in Malawi addressing this conclusion. A bit more of this information, as derived from the findings, would have helped link the evaluation use the conclusions to more effectively link the findings to the final recommendations.
<b>Question 16.</b>	Are logical and informative lessons learned identified? [N/A if lessons are not presented and not requested in ToR]		
	i Identified lessons stem logically from the findings, have wider applicability and relevance beyond the object of the evaluation.	Partially	Nine lessons are briefly described that cover relevant areas such as policy relevance, cross-sector collaboration, capacity building, community support and monitoring. While these respond broadly to the evaluation findings, it is not sufficiently clear that these add additional information that supports decision-makers beyond the JPGE, whether in UNICEF and its UN partners, or among their government counterparts.
	ii Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	Although the statements add somewhat to the evaluation conclusions and provide some useful insights into critical areas for future consideration, they are a bit short in detail and add little to what has already been said in the ER. While there does appear to have been a review of the conclusions and lessons as a joint section with mutually supporting elements that should be considered together, the sections do not sufficiently complement one another, add to the evaluation findings or work as stand-alone sections to which the evaluation users can usefully refer in order to build a quick and insightful picture of what the evaluation is trying to say.

SECTION G: RECOMMENDATIONS (weight 15%)		50%	Comments on Rating
<b>Question 17.</b> Are recommendations well grounded in the evaluation?			
i	Recommendations align with the evaluation purpose, are clearly formulated and logically derived from the findings and/or conclusions.	Partially	Twenty-four recommendations are described. Each is preceded by a short 'challenge' statement which is useful. It is also helpful that each of the recommendations is mapped back to relevant findings making it clear from what area of analysis the recommendations were derived. Different priority levels are allocated, alongside proposed timelines and responsibilities. Nevertheless, the decision to present the recommendations against the evaluation criteria does not help the reader relate the recommendation to the evaluation objectives and can lead to unnecessary overlaps. This leads to a tendency for the recommendations to be located at the level of individual actions rather than the strategic shifts required for a successful follow-up to the JPGE. For example, some of the recommendations under different criteria are related and could have been better clustered under a higher level recommendation where overall responsibility can be clearly seen. This is could be the case for the recommendations relating to disability (Recommendation E2 which falls under the efficiency criterion) and gender, outreach and inclusion (Recommendations EF 4,5,6,7,8,9 under effectiveness) which might have been presented together, allowing for greater clarity over the range and scope of individual agency leadership responsibilities and supporting roles.
ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Partially	Twenty-four recommendations is a long-list. To be clearly actionable, up to 7 would be more appropriate. Without this the recommendations can appear a little superficial. For example, the 6 current recommendations on coherence could be integrated into two, focusing on the strengthening of district and national and level coordination (currently recommendations C1&2, and C3,4,5&6 respectively, as well as EF10 under effectiveness). This would have supported a much clearer sense of who (among the multiple agencies and government bodies listed against each of the recommendations) should be held responsible for the overall implementation of the recommendation. Rather than simply providing top-line statements, the recommendations should pay greater attention to providing some level of guidance and support to users. For example, the second coherence recommendation responds to the challenge that, "Coordination platforms lacked formal accountability and sustainability mechanisms" by recommending the need to, "Introduce performance-based reviews for coordination platforms and assign district-level government leads to ensure accountability and continuity". Whilst relevant, it is doubtful that this action alone will address the challenge, or that the multiple stakeholders (many of them outside the government system) identified as responsible for implementing the recommendation are appropriate. Greater specificity and depth are required as to who exactly should do what.
iii	Process for developing the recommendations is described, and includes the Involvement of duty-bearers, as well as rights holders when feasible (or explanation given for why they were not involved).	Partially	The introduction to the evaluation recommendations in section 5.3 of the report provides a brief outline the two step process of developing the recommendations and acquiring feedback. However, it is unclear whether and how critical government stakeholders were included.
<b>Question 18.</b> Are recommendations clearly presented?			
i	Clear identification of groups or duty-bearers responsible for action for each recommendation (or clearly clustered group of recommendations). Clear prioritization and/or classification of recommendations to support use.	Partially	While duty-bearers are listed, it is difficult to understand who exactly should be responsible for overall leadership of individual recommendations (or sub-recommendations). The clustering of recommendations against the evaluation criteria is at times confusing, feels unnecessary, and leads to potential gaps in the review of oversight and follow-up that would have been possible through an alternative way of structuring around key strategic follow-up areas that relate directly to the evaluation questions.
<b>SECTION H: REPORT STRUCTURE AND PRESENTATION (weight 5%)</b>		<b>79%</b>	<b>Comments on Rating</b>
<b>Question 19.</b> Does the evaluation report include all relevant information?			
i	Opening pages include: Name of evaluated object, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents -including, as relevant, tables, graphs, figures, annexes; list of acronyms/abbreviations, page numbers.	Yes	The introduction is well presented and contains all relevant information as listed in the question.
ii	Annexes include: terms of reference, evaluation matrix, list of interviewees, results chain/ToC/logical framework (unless included in report body), list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology, information about the evaluator(s), etc.	Yes	The annexes include the evaluation terms of reference, framework (matrix), JPGE results framework, theory of change, bibliography, participants, methodology, national trends in education, and interview protocols and tools.
<b>Question 20.</b> Is the report logically structured?			
i	Structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles, well formatted).	Yes	The report is clearly presented and structured. All sections are appropriately numbered and formatted.
ii	Structure accords to UNICEF guidelines for evaluation reports: context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations.	Yes	All relevant sections are included in line with the UNICEF template and follow a clear logic order that is easy to follow.
<b>Question 21.</b> Is the report well presented?			
i	Report is of reasonable length; it does not exceed number of pages that may be specified in ToR.	Partially	At 100 pages the report is too long and could have been edited down significantly. This could have been achieved by presenting the bulk of the evaluation methodology in an expanded Annex 7, and reducing the findings narrative by presenting more of the analysis in tabulated summaries or figures, and using boxes to provide illustrative examples.
ii	Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Partially	While the report is reasonably well written and easy to understand, a few errors remain with some sentences left hanging or without the required punctuation. For example, in the methodology section and recommendationson (e.g. "C1a: Recommendation 4..."). A short spell-check, grammar review and edit would have been beneficial.
iii	Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labeled, and referenced in text.	Partially	While there is some good use of visual aids to break up the narrative, including under the findings for efficiency and impact, many do not include relevant sources. A wider use of tables and figures to summarise and present some of the qualitative as well as quantitative findings would have been helpful and may have helped reduce the length of the report.
<b>SECTION I: EVALUATION PRINCIPLES (weight 10%)</b>		<b>86%</b>	<b>Comments on Rating</b>

<b>Question 22.</b>	Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity?		
	i Reference and use of rights-based framework, and/or CRC, CCC, CEDAW and/or other rights related benchmarks in the design of the evaluation.	Partially	The report includes clear reference to child rights and rights-based benchmarks for such as the Convention on the Rights of the Child, and references education and sexual and reproductive rights, it does not provide references to supporting documents or include adherence to the Convention on the Rights of Persons with Disabilities.
	ii Clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation (for example, a reference group is established, stakeholders are involved as informants or in data gathering).	Yes	All relevant stakeholders are identified and their levels of participation are described through the description of the evaluation methodology. This included the engagement of UNICEF, UNFPA, UNESCO, and national and district government staff and external stakeholders such as children, community members, parents and youth.
	iii Language is empowering and inclusive, avoiding gender, heterosexual, age, cultural and religious bias, among others; use terminology of rights holders and duty bearers; data is disaggregated by marginalized group; differential results are assessed (distribution of results across different groups).	Partially	The evaluation does not show evidence of any gender, cultural, age, disability or other bias. The report is transparent as to the evaluation limitations. A robust presentation of sex, age and disability disaggregated data is provided. The report does not, however, use the terminology of rights holders and duty bearers.
	iv Evaluation assesses the extent to which the implementation of the intervention addresses child rights and Leave No-one Behind (gender and other excluded and marginalized groups). It is disability inclusive.	Yes	The evaluation explores UNICEF adherence to leave-no one behind principles through the lens of girls' access and inclusion in education, adolescent girls' support, and disability-inclusion. Each is strongly integrated in the evaluation subject, questions and methodology, as well as in the evaluation's collection of evidence, analysis of findings and presentation of final recommendations.
<b>Question 23.</b>	Does the evaluation meet UN SWAP evaluation performance indicators? (Note: this question will be rated according to UN SWAP standards with detail provided below)	9	
	i GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Fully integrated	Gender is integrated in the evaluation subject and scope, the methodology, questions, evaluation matrix and analysis of findings. Analysis of the relevance criterion includes examination of the extent to which the JPGE supported girls' participation in education and the wrap-around support provided both in and out of schools.
	ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Fully integrated	The methodology provides a strong example of gender analysis in the data collection and analysis tools which used mixed quantitative and qualitative methods triangulated across multiple stakeholders.
	iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Fully integrated	Gender is a clear and dominant part of the evaluation background, methodology, analysis of findings, presentation of conclusions and lessons learnt, and formulation of the evaluation recommendations.

#### SWAP Rating Guidance

<p><b>i GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.</b></p> <p>a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?</p> <p>b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?</p> <p>c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?</p> <p>d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?</p>
<p><b>ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.</b></p> <p>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex?</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?</p> <p>d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?</p>
<p><b>iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.</b></p> <p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality?</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable?</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described?</p> <p>d. Does the evaluation report provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area?</p>